



# Artificial intelligence and identity: the rise of the statistical individual

Jens Christian Bjerring<sup>1</sup> · Jacob Busch<sup>1</sup>

Received: 10 August 2023 / Accepted: 11 January 2024  
© The Author(s) 2024

## Abstract

Algorithms are used across a wide range of societal sectors such as banking, administration, and healthcare to make predictions that impact on our lives. While the predictions can be incredibly accurate about our present and future behavior, there is an important question about how these algorithms in fact represent human identity. In this paper, we explore this question and argue that machine learning algorithms represent human identity in terms of what we shall call the *statistical individual*. This statisticalized representation of individuals, we shall argue, differs significantly from our ordinary conception of human identity, which is tightly intertwined with considerations about biological, psychological, and narrative continuity—as witnessed by our most well-established philosophical views on personal identity. Indeed, algorithmic representations of individuals give no special attention to biological, psychological, and narrative continuity and instead rely on predictive properties that significantly exceed and diverge from those that we would ordinarily take to be relevant for questions about how we are.

**Keywords** Identity · Algorithms · Predictive properties · Statistical individual · Narrative continuity · Representations of humans

## 1 Introduction

In the digital era, our interactions leave behind a rich tapestry of data. Every click, every purchase, every “like”, every search query is recorded, and each contributes to a digital footprint of us as individuals with specific preferences, habits, interests, and behaviors. In this sense, they are a reflection of our identity, but through the lens of our digital interactions. Machine learning algorithms, which are central to many of today’s technological systems, are designed to make sense of these vast amounts of data. They process the digital footprints we leave behind and look for patterns and correlations among the different data points. For example, an algorithm might learn that people who buy product A often also buy product B, or people who watch movie M also typically watch movie N. As such, algorithms essentially translate our raw digital footprints into a more structured format that can be analyzed more easily.

One common way of structuring these data is to represent each individual as a vector in a high-dimensional space. Intuitively, we can think of each dimension in this space as corresponding to a different property, attribute or behavior. For example, one dimension might correspond to the property of liking science fiction movies, while another might correspond to a property that represents how frequently you shop online. The specific value that an individual has on each dimension—the individual’s coordinates in this space, so to speak—forms a unique profile that represents the individual. By analyzing such profiles, algorithms can make predictions about people’s present and future behavior. If an individual’s profile is similar to a group of people who have shown a particular behavior—say, they all enjoyed a specific movie—then the algorithm might predict that the individual will also enjoy that movie. This predictive capability is what powers recommendation and classification systems, targeted advertising, and many other modern technologies.

However, while these algorithmic predictions can be incredibly accurate, there is an important question about how these algorithms in fact represent human identity. This is the question that we want to address in this paper. More specifically, we will argue that algorithms represent human identity in terms of what we shall call the *statistical individual*. This statisticalized

---

✉ Jens Christian Bjerring  
filjcb@cas.au.dk

Jacob Busch  
filjab@cas.au.dk

<sup>1</sup> Department of Philosophy, Aarhus University, Jens Chr. Skous Vej 7, 8000 Aarhus C, Denmark

representation of individuals, as we shall see, differs significantly from our ordinary conception of human identity.

Ordinarily, we take it, our representations of human identity are tightly intertwined with considerations about *biological*, *psychological*, and *narrative* continuity—as witnessed by the fact that our most well-established philosophical views on personal identity appeal to those very considerations. Biological continuity roots our identity in our biology, capturing the thought that we remain the same individuals although we age and undergo various other physical changes during a lifetime. Psychological continuity weaves together our past, present, and future experiences, beliefs, and memories, capturing the thought that a certain degree of psychological connectedness between our different temporal parts matters for our sense of being a unique person. Narrative continuity grounds our identity in the life stories or narratives that we construct about ourselves, capturing the thought that who we are is to a large extent informed by how we—as socially and culturally embedded people—interpret and give meaning to our past, present, and future experiences. Importantly, the properties that characterize biological, psychological, and narrative continuity are properties that we can readily recognize as being relevant to questions about our identities. Such properties, intuitively speaking, directly relate to and reflect the individuals we are and represent ourselves as being.

Algorithmic representations of individuals, however, do not prioritize these facets of identity: they give no special attention to biological, psychological, and narrative properties. As such, we argue, algorithmic representations fail to capture central aspects of our ordinary representations of human identity. Indeed, we shall argue, algorithms make predictions about us by appealing to properties that we struggle to square with our own representations of who we are.<sup>1</sup> In contrast to the properties, which characterize biological, psychological, and narrative continuity, algorithms make use of predictive properties that do not directly relate to and reflect the individuals we understand and represent ourselves as being. Or, as Milano et al. remark in a different context, “even if users could access the content of the [algorithmic] model, they would not be able to interpret it and connect it with their lived experiences in a meaningful way” (Milano et al. 2020, p. 962).

<sup>1</sup> So if we are correct, then even if some version of *connectionism* is true—in which case statistics, predictions, and pattern recognition are at the very root of not just machine learning systems, but also human cognition—humans and algorithms will represent individuals very differently. There are intriguing questions about the causes of such differences in representation—especially if the underlying statistical architectures are similar, as connectionism seems to imply—but for the purposes of this paper, we are happy just to argue for the claim that human and algorithmic representations of individuals *are* significantly different. For further discussion of these and related matters, see Humphreys and Yoshihisa (2002) and Sutton (1998). Thanks to an anonymous referee for raising these issues.

To the best of our knowledge, there is no current research that gives a detailed analysis of algorithmic representations of human identity and how it may conflict with our own representations of who we are. Other pertinent questions surrounding the multifaceted notion of human identity have of course been addressed in AI contexts. Questions about human identity are interwoven with questions about the human body. Lagerkvist et al. (2022) suggest that algorithmic biometrics will result in a novel type of objectification of the human body, highlighting the potential ethical and existential ramifications that such an objectification may cause. In that vein, Babushkina and Votsis (2022) have shown how close human–machine interactions and pairings—where algorithms act as cognitive extenders or constituents of an extended mind—can impact various social practices involving human agency; see also Søe and Mai (2022). Intuitively, as proponents of the extended mind thesis noted, tampering with your iPhone may mean tampering with your memories and desires; see Clark and Chalmers (1998) and Pedersen and Bjerring (2022). Questions about human identity are also tightly interwoven with questions about autonomy and authenticity. Recent studies show how algorithms—notably AI-powered recommender systems—may affect autonomy in ethically problematic ways; see del Valle and Lara (2023) and Milano et al. (2020). The observations that algorithms and AI may have disruptive consequences for our understanding of the human body, the relationship between human and technology, and autonomy align well with the types of conclusions that we reach in this paper. For once we have unpacked the algorithmic representation of us as statistical individuals, it will not be hard to appreciate how uniquely different their representations of us are compared to how we ordinarily think of ourselves as individuals.

For the purposes of this paper, when we talk about algorithms, we talk about the class of machine learning algorithms. So when we say that algorithms represent us as statistical individuals, we mean that machine learning algorithms represent us as such. Abstract statistical methods, as we shall see more clearly in Sect. 3, underpin most machine learning algorithms. But they may of course differ significantly in both their architectures and methods. For instance, convolutional networks, recurrent networks, and transformers are all types of neural networks that rely on statistical methods for representing data, but their goals and architectures are quite different. Convolutional networks use convolutional layers to detect patterns in data, recurrent networks utilize feedback loops to capture sequential dependencies in data, and transformers appeal to attention mechanisms to uncover global dependencies in data. Similarly, support vector machines, logistic regression, and linear regression are all statistical algorithms, but they are associated with quite different statistical methods.

Support vector machines attempt to separate data points into two classes—by finding a hyperplane that maximizes the distance to the nearest data point from each class—that are then used for classification and regression analysis, logistic regression aims to predict the probability of a binary outcome by fitting a logistic function to the data, and linear regression tries to predict a continuous output using linear equations to model relations between the dependent variable and one or more independent variables.<sup>2</sup>

While we thus recognize that machine learning algorithms can be different in many ways, their abstract statistical nature, as we shall see in Sect. 3, make them similar enough for our purposes to be classified as one type and as giving rise to a statistical representation of individuals. But of course not all algorithms are statistical algorithms. Many “good old AI” symbolic systems are still widely used today, and even machine learning models often incorporate techniques from classical rule-based knowledge-systems and explicitly programmed logical systems.<sup>3</sup> While these types of symbolic AI systems do not give rise to a statistical individual—as they are not statistical in nature—it is natural to speculate whether they could give rise to something like a *symbolic system individual*.<sup>4</sup> After all, insofar as we can use symbolic systems to store, retrieve, manipulate, and reason about individuals and their properties, it is plausible, as far as we can tell, to think that such symbolic AI systems may also yield representations of individuals that differ from how we ordinarily think of ourselves as individuals. Although we shall not pursue this thought further in this paper, we are sympathetic to the idea that we can have a whole spectrum of distinct algorithmic representations of human identity. As we shall see more clearly in Sect. 4, we focus on machine learning algorithms because their use of high-dimensional vector representations of individuals result in representations that can come dramatically apart from the types of representations that we find in the philosophical literature and through introspection. But this is not to say that similar conclusions cannot be reached by appeal to other types of AI systems.

Here is how we proceed. In Sect. 2, we briefly lay out the central tenets of the biological, psychological, and narrative views on what constitutes a person; we will spend a bit more time on the narrative view as it is the less well-known view and as it will play a more central role in our later discussions. Obviously, it is not our aim to present nor criticize these different views on identity in any detail. Instead, it is

to provide sufficient context to highlight the distinction between algorithmic representations of individuals and the complex blend of biological, psychological, and narrative elements that constitute our ordinary understanding of individuals. In Sect. 3, we introduce the concept of the statistical individual and show that it does not give any special weight to biological, psychological, and narrative properties. In Sect. 4, we elaborate on the unique properties of the statistical individual and argue that these properties need not directly relate to nor directly reflect the individuals who are affected by the relevant algorithmic predictions. In Sect. 5, we offer a few concluding remarks.

## 2 Views on identity

What makes a specific individual that very individual? While many different answers to this question have been aired in the philosophical literature, we will focus on the three, arguably, most prominent answers: the biological, the psychological, and the narrative view. In our opinion, as mentioned, these views capture central aspects of our ordinary understanding of human identity, and they will constitute the backdrop against which we can appreciate how unique the algorithmic representations of humans are.

### 2.1 Biological and psychological views

When we ask what makes a specific individual that very individual, we typically place central importance on biological and psychological properties. The biological and psychological views in philosophy erect theories of personal identities from these premises.<sup>5</sup> According to a *biological* view, an individual  $X$  at time  $t_1$  is the same individual as individual  $Y$  at  $t_2$  just in case  $Y$ 's biological organism is continuous with  $X$ 's biological organism.<sup>6</sup> Of course, in appealing to facts about biological organisms, advocates of the biological view are not restricted to identifying individuals over time by merely looking at their bodily properties. We can appeal to whatever properties we deem relevant for characterizing biological animals like us. Yet, as a matter of fact, clearly observable bodily facts are often what we appeal to when determining whether an individual in the past is identical to an individual in the present. By contrast, the *psychological* view on personal identity holds that an individual  $X$  at time  $t_1$  is the same individual as individual  $Y$  at  $t_2$  just in case  $Y$  is sufficiently psychologically continuous with  $X$ .<sup>7</sup> Roughly, for the individual  $X$  at  $t_1$  and the individual  $Y$  at  $t_2$  to count as

<sup>2</sup> For more information on the various different types of machine learning algorithms, see for instance Ayodele (2010) and Sarker (2021).

<sup>3</sup> For more on the symbolic AI paradigm and its relation to machine learning systems, see for instance Garnelo and Shanahan (2019).

<sup>4</sup> Thanks to an anonymous referee for raising this idea.

<sup>5</sup> For purposes of stating the biological and psychological views of identity, we rely on the formulations in Shoemaker (2008).

<sup>6</sup> See, for instance, Olson (1999) for an elaboration of this view.

<sup>7</sup> See, for instance, Shoemaker (1963) for an elaboration of this view.

the same individual,  $X$  and  $Y$  must share enough—whatever that amounts to precisely—of the same memories, desires, beliefs, and psychological dispositions.

Typically, psychological continuity goes with biological continuity, but when it does not, thought examples seem to suggest that intuitions about personal identity follow facts about psychological continuity. Peter has committed a series of brutal murders, whereas Andrea has never even received a speeding ticket. Doctors transfer Peter's brain into Andrea's debrained body in such a way that the individual with Andrea's body is psychologically continuous with Peter immediately before the surgery. As a result, the individual who has Andrea's body after the surgery has all the memories, all the desires and hopes, and all the beliefs and psychological dispositions that Peter had before the surgery. In a case like this, most would intuit that the individual in Andrea's body is still Peter due to the psychological continuity between the two entities. So, at least *prima facie*, such body transplant intuitions seem to speak in favor of the psychological view.<sup>8</sup>

Abstracting away from their differences and focusing on their similarities, we can understand both the biological and psychological views as capturing the familiar thought that an individual's 'life trajectory' moves from the *past* towards the *future*. Indeed, on these views, there are clear causal connections between the past, present, and future selves: who we are now is directly causally related to the biological or psychological organisms that we used to be, and who we will be in the future is directly causally related to the biological or psychological organisms that we are now. Moreover, biological and psychological properties are properties that people, intuitively speaking, are willing to grant as being *directly related* to or as *directly reflecting* who they are as individuals. Suppose Peter has been recorded by a CTR camera for stealing an apple at some point in the past. To hold Peter responsible now for stealing the apple, Peter must be identical to the individual on the CTR image. Imagine that all the CTR images of the individual stealing the apple include a butterfly in the background. It would be odd, to say the least, if we appealed to this butterfly in adjudicating questions about whether Peter is identical to the person on the CTR camera. Rather, when it comes to identifying individuals over time, the sensible properties to appeal to are somehow properties that directly relate to or reflect an individual. While we will be more precise in Sect. 3, these properties are intuitively properties that individuals would be willing to accept as being relevant to questions about their identity. Whereas butterflies are irrelevant for such questions, typical biological and psychological properties are not.

But psychological and biological continuity is not all that matters for our ordinary conception of what makes up a person.<sup>9</sup> Life stories or narratives, as we shall see now, also play a central role.

## 2.2 The narrative view

According to a *narrative* view, what makes a specific individual that individual is in some sense story-like or narrated. There are different ways of cashing out this idea, but at the core is the thought that questions about who we are can only be answered in the context of a narrative.

When we view questions about identity through the lens of life stories, an individual's actions and

experiences must be actively unified, must be gathered together into the life of one narrative ego by virtue of a story the subject tells that *weaves* them together, giving them a kind of coherence and intelligibility they wouldn't otherwise have had. This is how the various experiences and events come to have any real meaning at all—rather than being merely isolated events—by being part of a larger story that relates them to one another within the context of one life. (Shoemaker 2021.)

On the narrative view, then, we make sense of an individual's actions and experiences by embedding them in a life story. Consider again the CTR camera catching Peter putting an apple in his pocket. We can characterize Peter's behavior in different ways. We might say that Peter was engaged in theft, that Peter thought the apples were free to take, or that Peter was rescuing his children from starvation. On the narrative view, which characterization we choose as the most appropriate of Peter's behavior depends on the ongoing story or narrative of Peter's life. If the story weaves Peter's life together with the near impossibility of survival in his world, we might characterize the event of Peter putting the apple in his pockets in terms of survival rather than in terms of theft.

On the narrative view, individual actions—individual scenes from a life—thus only gain proper meaning when they are embedded in an individual's life story. But not just any life story goes. It is hard, if not impossible to make sense of life stories that do not display an appropriate level of coherence. For instance, I cannot coherently describe some past event in my life with a statement such as "I used to love being a bachelor, in particular when my wife was still around", unless of course I am intending a pun or do

<sup>8</sup> Advocates of the biological view have of course given sophisticated answers to cases of brain transfers, but the details need not worry us here; for discussion, see for instance Parfit (1986).

<sup>9</sup> For more information on how the narrative view complements and extends the psychological and narrative views, see for instance DeGrazia (2005), Schechtman (1996, 2011), Shoemaker (2008), and Shoemaker (2021).

not mean with the English language what other people do. From an individual perspective, a sense of coherence is relevant to establishing a sense of identity over time. For a person to identify with some character attribute, for example, there has to be a sense of coherence between that attribute and the individual's own narrative about their character. But internal coherence is not all that matters. There also needs to be some level of correspondence between the content of an individual's life story and the ongoings in the world around the individual. Absent severe psychological illness, it is difficult to construct a meaningful narrative in which an individual makes sense of his experiences in the morning through a narrative based on events surrounding Julius Caesar's crossing the Rubicon River, and in the evening through a narrative based on events surrounding Marilyn Monroe's marrying James Dougherty. Rather, meaningful narratives are informed and constrained by what is going on in the world.

Obviously, it is not easy to demarcate precisely how the world and other individuals in it constrain the life narratives of individuals. But it is clear that life stories are highly affected by the social embedding of the individual whose life story is being narrated. As Schechtman puts it:

We are composing the stories of our lives in a vacuum, but in a world where there are others with their own stories about themselves and about us. [...] Both because our narratives must make reference to the stories available to us from the traditions in which we find ourselves and because they must interact with the realities of the world in which we live and the narratives of others, our narratives must be understood as embedded in a world of other selves. (Schechtman 2011, p. 405.)

Indeed, it is plausibly the case that other individual narratives are *constitutive* of the life stories that individuals weave. If an individual cannot coherently narrate a story in which he has powers equal to an omnipotent god, this is in part because individuals around him do not treat him as such a god: they do not move when he intends them to move, and they do not stop driving their cars when he wants them to. If the coherence and intelligibility of a life story in this sense depends on whether other individuals collaborate it, life stories are constitutively dependent on the narratives of others. In this sense, the narrative view differs from both the psychological and biological views: it expands, intuitively, the set of properties that we deem relevant for answering questions about who we are.

In the ordinary run of things, we can, as above, understand the story of an individual's life as flowing from the past towards the future. Individuals use past elements of the narrative to make sense of current events. Peter, for instance, might regard his driving an expensive car as justified because of all the hard work that he put into his university

studies years back. Similarly, individuals use current and past elements of the narrative to anticipate future events. When Peter, for instance, makes sense of his countless runs in the cold autumn rain, he does so by embedding them into a larger narrative that extends into a future where he is slim and fit for summer. So the key move, for the narrative view, concerns

[...] the claim that narratives have a kind of diachronic holism that psychological continuity as it is understood in psychological continuity theories does not. While psychological continuity is defined as a relation between independently definable time-slices, in a narrative the parts exist in the form they do only as abstractions from the whole, and so the whole is, in an important sense, prior to the parts. (Schechtman 2014, p. 100.)

This idea of a *temporally extended narrative self*, as we may call it, captures important aspects of the ordinary conception of identity.

By grounding identity in temporally extended narratives, the narrative view thus entails that an individual's life trajectory—moving from the past towards the future—is not merely a chain of biologically or psychologically linked events. Instead, it is woven together through the overarching narrative that the individual constructs, where each life event gets interpreted and contextualized within the broader story of their life. That is, who we are now is connected to who we used to be through a narrative, and who we will be in the future is connected through a narrative to who we are now. But while the narrative view does not posit a causal, biological, or psychological link between life events in the past, present, and future, it does require these events to be connected by a coherent and unifying life story. In this sense, the narrative view imposes constraints on the possible life trajectories of an individual with a particular past—the constraints are just not of a causal, biological, or psychological kind. To witness, we expect individuals to act in the present with care for their future selves. But not necessarily with care for any old future self. For while there are many possible future time-slices of an individual, the individual typically only cares about those future time-slices that are, properly understood, extensions of the specific story that narrates his or her life. Suppose 60-year-old Peter expects a future as an active hiker in his retirement. As such, Peter is concerned only with future time-slices of himself that satisfy certain properties such as being alive, being sane, being minimally healthy, being an outdoor lover, and so on.<sup>10</sup> That is, the set

<sup>10</sup> In a sense, just like Kripke's Humphrey cares about whether *he*—and not some counterpart of his in some distinct possible world—will win the next election, so Peter cares about whether *he*—as an extended narrative individual—will be able to live the life that he

of future time-slices that are of concern to Peter are those that match the temporally extended narrative that makes up Peter's life. In this sense, individuals act on the assumption that there is an identifiable future self, which their past and present actions partially narrate and help create. So even if narrative continuity constrains the set of possible life trajectories in a significantly looser way than biological and psychological continuity do, it nevertheless draws the clear contours of a life that moves in a past-to-future direction and that the individual recognizes as involving him or her as a central actor.

While the narrative view thus extends our understanding of identity beyond the class of biological and psychological properties, the properties of life narratives can still straightforwardly count as properties, which, in the vocabulary from above, directly relate to or directly reflect who we are as individuals. After all, the properties characterizing an individual's life story do so precisely because they are expressed in the individual's own narrative of his or her life.<sup>11</sup>

### 3 Algorithmic predictions and the statistical individual

In conjunction, then, we can see how the biological, psychological, and narrative views capture important aspects of our ordinary representations of human identity. Importantly, the properties that these different views rely on are properties that individuals can readily recognize as being relevant for questions about who they are—they do, in the vocabulary above, directly relate to individuals, or, in the vocabulary below, count as self-relatable properties for individuals. Yet, as we shall argue next, algorithms represent human identity through a set of properties that need not place any obvious centrality on biology, psychology, or narrative coherence.

As mentioned in the introduction, when we talk about algorithms in this paper, we talk about machine learning algorithms. Despite their differences, such algorithms are able to learn patterns in vast amounts of data, and when used for predictive purposes, they can use these learned patterns to determine the likelihood of individuals having certain properties. In a medical context, a machine learning algorithm might have learned a correlation between complex patterns of patient attributes and a certain risk of developing testicular cancer. When it comes to a specific patient with specific values for the relevant patient attributes, the algorithm can then give the patient a specific risk score

of developing testicular cancer.<sup>12</sup> In a financial context, a machine learning algorithm may have learned how to correlate certain facts about individuals—facts such as age, residential location, annual income, and stock holding—with a financial risk profile.<sup>13</sup> In this sense, machine learning algorithms are thus trained—typically in a supervised manner—to find correlations in massive datasets between sets of relevant input parameters and specific output parameters. Once the algorithm has successfully learned to accurately correlate the input parameters with the output parameters, the algorithm can be utilized to make predictions about individuals.

Since machine learning algorithms trade in statistical correlations, we can even more abstractly think of predictive algorithms as giving us a probabilistic estimate of an outcome based on a particular set of values for the algorithm's input parameters. In a banking context, based on information about facts such as your age, residential location, annual income, and stock holding, an algorithm may give you a certain probability of defaulting on a specific loan. To make as precise predictions about individuals as possible, it is generally useful to have many input parameters to feed the algorithms. Of course, we do not want to risk overfitting the training data, but the real wonders of machine learning algorithms—and, in particular, deep neural nets—concern the large number of parameters that they can shuffle. By having more input parameters to characterize individuals, we can get closer to finding a set of properties that uniquely characterize these individuals. That is, a machine learning algorithm with 100 relevant input parameters can, intuitively speaking, give a more fine-grained representation of an individual than a similar model with only 10 relevant input parameters: knowing your annual income *and* your stock holding gives me a more fine-grained representation of you as a financial individual than merely knowing your annual income.

We can think of an input parameter space for a machine learning algorithm as serving as a reference class from which the algorithm can draw inferences about individuals based on the patterns it has learned. For example, if the input parameters in a credit scoring algorithm include features such as age, residential location, annual income, and stock holding, the relevant reference class would be the set of individuals represented by the various combinations of these features. The algorithm would then have learned patterns within this reference class to make predictions about creditworthiness for new individuals with similar values for the input parameters. In this way, by having richer relevant input parameter spaces, we can sort individuals into more fine-grained

Footnote 10 (Continued)

wants to in the future (Kripke 1972).

<sup>11</sup> For more on the relationship between the narrative and the biological and psychological views, see Schectman (1996), pp. 130–136.

<sup>12</sup> For more information, see for instance Bjerring and Busch (2021), Jiang et al. (2017), and Liu et al. (2019).

<sup>13</sup> For more information, see for instance Leo et al. (2019).

reference classes and, derivatively, get closer to finding sets of properties that uniquely represent these individuals.

Based on the particular reference class that an individual belongs to, many predictive machine learning algorithms are thus used to make inferences about what further properties that individual will likely have. For instance, an algorithm may infer that there is a high probability that individual X in the future will default on his loan given that X statistically belongs to a group of people with these-and-these specific properties. In this sense, when algorithms are used to make predictions about individuals, they represent individuals as *statistical individuals*. Put crudely—we will be more precise soon—the properties that define the statistical individual are those properties that define the reference class to which the individual belongs and those that the algorithm can statistically infer belong to the individual in virtue of the individual's membership in the particular reference class.

To illustrate in more detail what this statisticalized representation of an individual is, consider a machine learning algorithm that gives individual X a credit score, and suppose the score is too low for the bank utilizing the algorithm to issue X a loan. In this type of case, the bank denies X a financial opportunity based on a probabilistic estimate that X will engage in a type of behavior—namely bad credit defaulting—in the future. While the algorithmic prediction relies on some conception of X that extends into the future, it is easy to see that the relevant conception need not rely on any biological or psychological continuity between X and the statistically individuated future time-slice of X who defaults on the loan. Algorithmic representations of individuality thus extend into the future, but the relevant future selves are those that are *statistically* related to the individual X now. Yet, being statistically related to an individual now is not the same as being biologically or psychologically related to the individual. After all, even if X belongs to a reference class whose members have a high risk of making a bad default on a loan, we are still only dealing with a probabilistic risk estimate: the individual Y in the future who is biologically and psychologically continuous with X might in fact *not* default on the loan.

Accordingly, there is no guarantee that the statistical representation of individuals respects any non-trivial biological or psychological characterization of individuals—it may, of course, if the relevant algorithm's input parameter space includes various biological and psychological features, but it *need not*. Given that algorithmic predictions are often accurate, the properties of a biologically and psychologically continuous individual may in fact frequently align with those of the corresponding statistical individual. But conceptually, they are very distinct.

Something similar goes with respect to the narrative view. Consider again individual X and suppose his loan application was denied because of some specific set of financial problems

in his near past. While the credit scoring algorithm uses these properties of X to predict that X will eventually default on his loan, the narrative view might tell a completely different story about how these events in X's past project into his future. On the narrative view, for instance, there might be very good reasons for the financial problems that X has endured in the past. Indeed, the numerous overdrafts and overdue invoice reminders might form part of a temporally extended narrative that tells the story about an individual who has learned about finances the hard way—not unlike the ex-addict who takes his past experiences and actions as formative for the current person he is. Where the algorithm punishes X for the financially irresponsible past behaviors, X himself understands these behaviors as necessary stepping stones for the financially responsible individual that he is now and intends to be in the future. As such, the interpretation of X's past behaviors inside a narrative can draw the contours of a particular future individual that is very different from the statistical individual that the algorithm represents based on the same set of past behaviors. So, again, there is no guarantee that the statistical representation of individuals bears any interesting relation to the temporally extended narrative self. Of course, the algorithmic representation can agree with the narrative view that we must go beyond biological and psychological properties—indeed, beyond any sort of causal continuity—to represent human identity. But conceptually, they again remain very different.

Accordingly, when it comes to machine learning algorithmic predictions, there need be no biological, psychological or narrative continuity between the individual affected by the algorithm prediction and the statistical individual represented by the algorithm for purposes of said prediction. Algorithms place no special focus on biological, psychological, or narrative continuity between different temporal stages of an individual. Rather, the focus is on the statistical relations that obtain between the properties in the reference class to which the individual can be said to belong.

#### 4 The statistical individual and its unique properties

But the contrast between our ordinary representations of human identity and the statisticalized representations made by algorithms gets even starker. As we saw, the properties that characterize a statisticalized representation of an individual are those properties that define the reference class to which the individual belongs and those properties that the algorithm statistically can infer belong—with a certain probability—to the individual in virtue of its membership in the particular reference class. Quite often, we are familiar with the properties or parameters that define specific reference classes. In typical cases of credit scoring, for instance, the

reference class will include a set of properties that capture various aspects of an individual's payment history. For the algorithm, that is, there will be input parameters measuring how many bills were overdue in January, how many were overdue in February, and so on, and there will be parameters that measure the amount of bills paid over the last, say, 12 months. We can expect some of these parameters to be dependent on each other. For instance, if there is an increase in the total number of bills that need to be paid, then there is likely an increase in the number of bills whose payment is overdue. In addition to such parameters, which intuitively tell us something about the financial reliability of the individual as a borrower, there are also parameters for familiar demographic data such as sex, education, age, marital status, employment history, time and place of birth, home address, and so on.

There is a clear sense in which this kind of information reflects properties that we take to be rather unique to the individual in question. Looking at the narrative view, for instance, facts about financial reliability, age, marital status, education, and employment history will likely constitute a major part of an individual's life story. As such, it makes sense to see these facts as directly relating to or reflecting the individual: they inform to a large extent the temporally extended narrative. Whether or not we take parameters such as education, employment history, and spending behavior to have the same significance for credit scoring as algorithms do, we can nonetheless appreciate why these parameters are used for making predictions about our financial situation. After all, it makes sense to tailor our loan abilities to our income and spending behavior, and we can all agree that we can somewhat affect the values for these parameters by the choices and decisions we make in life.

However, machine learning algorithms often make use of input parameters that reflect properties that we struggle to square with our own representations of who we are.<sup>14</sup> Consider an example involving algorithmic predictions that assist social workers in devising job training courses for unemployed people. In particular, let us consider an algorithm called *ASTA* whose job is to make an automatic assessment of unemployed Danish citizens' risk of long-term unemployment.<sup>15</sup> While many of *ASTA*'s 50 input

parameters reflect properties that we can intuitively think of as being directly relevant for an individual's job readiness, the following parameters clearly do not:

- whether the citizen allows the job placement center to send him or her text messages;
- whether the citizen allows the job placement center to send him or her emails;
- whether the citizen lives in an apartment;
- the number of different social workers involved with the citizen's case at the job placement center;
- the average time of day for held meetings between the job placement center and the citizen; and
- the longitude of the citizen's residence.

To be sure, we have already learned from the narrative view that we should extend our representations of human identity beyond the realm of the biological and psychological. Yet, an individual would be hard pressed—at least pre-theoretically—to make sense of why the parameters above should form part of a representation of him as an individual facing a certain risk of long-term unemployment. Intuitively, the *ASTA* parameters above just seem unrelated or unimportant to our ordinary understanding of which factors contribute to the risk of enduring long-term unemployment. That is, the *ASTA* parameters do not reflect properties that we would ordinarily accept as being relevant to questions about our occupational or professional identities.

To motivate this thought further, return to the narrative view and consider the parts of an individual's life story that concern their professional identity at some point in their life. While that part of their life story will plausibly include facts about properties such as level of education, age and past achievements—properties that readily reflect aspects of our professional identity—it will likely not contain information pertaining to the *ASTA* parameters above. Intuitively, when it comes to understanding the prospects of facing long-term unemployment, only few life stories will be narrated in terms of facts involving, say, the type of communication with the job placement center and the number of social workers involved in a case. Put differently, it is very likely only rarely the case that individuals will understand employability-related events in the past or anticipate employability-related events in the future through narratives that mention such facts. For such facts are simply not relevant for their professional identity: they do not reflect properties that we

<sup>14</sup> Refer to de Vries (2010), Milano et al. (2020), and Søe (2021) for discussions that touch upon somewhat related topics—although in very different vocabularies and settings from ours.

<sup>15</sup> For more on the *ASTA* algorithm and its properties, see Petersen et al. (2021) and Ammitzbøll Flügge et al. (2022). The context of the employment of *ASTA* is the “unemployment policy in Denmark and the use of AI and algorithms in the public sector more broadly. From the unemployed individual's perspective, Danish job placement is centered around the web application Jobnet. To receive unemployment benefits, an individual must register their unemployed status in

Footnote 15 (Continued)

Jobnet and participate in regular meetings with a caseworker. It is the responsibility of the individual to book meetings with the caseworker using the Jobnet portal and to update their ‘job log,’ providing the caseworker with an overview of the individual's job search activities.” (Seidelin et al. (2022), p. 2.) Note, though, that the *ASTA* algorithm is no longer in use.



would ordinarily understand as being relevant to questions about our employability.

Of course, we can reasonably speculate why parameters involving text based communication and average meeting times during the day can play a role in estimating length of time spent in unemployment. For instance, if an individual does not text or email, then it is likely more difficult for the individual to fit into a highly digitalized workforce, and if an individual never schedules meetings before noon, then this might suggest that the individual has unconventional sleeping patterns that make it hard for them to fit into a normal working culture. Yet, absent a more detailed input parameter space, such reasonable speculations remain but speculations. For it is not difficult to imagine many distinct explanations of why individuals cannot be reached by text or email. At the time of inputting their data to the ASTA algorithm, for instance, the individual might be between phones, or his email might have been hacked. The input parameters do not reveal such cases, and therefore their values might give us the wrong expectations about the job-preparedness of the individual in question.

Accordingly, high-dimensional vector representations of humans can come dramatically apart from the kinds of representations that we are ordinarily willing to accept as being relevant to us as individuals. Indeed, as machine learning algorithms grow more complex and ingest increasing amounts of higher dimensional data, it becomes more likely that many of their input parameters will represent features that are radically different from the types of features that we typically take be of relevance to questions about our identity.

Here is another example to illustrate this kind of observation. Consider again algorithms that assign individuals a credit score. While these algorithms typically emphasize financial history, including FICO parameters, they increasingly incorporate broader data sources—spanning cell phone usage to web browsing and social network interactions.<sup>16</sup> Companies like LenddoEFL, for instance, employ “alternative data sources including mobile phone, digital footprint, behavioral, and psychometric to assess the credit risk of anyone” (LenddoEFL n.d.). Although the specifics of such non-standard or alternative data remain proprietary, it is not hard to imagine that the algorithm will find useful predictive parameters in this data which, however, would only have little meaning in connection with our ordinary representations of who we are as financial individuals.

Take mobile phone data. These data might include information about call-detail records. Call-detail records are useful for constructing a rather detailed social network, which can be used to increase the predictive accuracy of credit scores. As studies by Kharif (2016) and Óskarsdóttir et al. (2019) suggest, machine learning algorithms might predict

an individual’s financial behavior in part based on the sheer volume of phone calls and text message interactions that the individual is involved in. Yet, for reasons of privacy, the algorithms are not looking at the *contents* of these calls and text messages. As such, while the sheer volume of phone calls and text messages might serve the algorithm’s predictive purposes, it is clear why we may struggle to square the properties that it uses for these predictions with our representations of ourselves as financial individuals. After all, why should properties reflecting volume of phone calls and text message interactions say anything important about me as a financial individual? Indeed, if I was denied a loan in part based on information about the total volume of phone calls and text message interactions that I have been involved in, I may reasonably question *why* this type of behavior should be relevant when it comes to assessing me as a financial individual.

Moreover, it is fairly well established that we can use cell and smartphone behavior as indicators of various personality traits.<sup>17</sup> When it comes to financial behavior, the personality trait of delayed gratification has been correlated with an increase in credit score, and delayed gratification has, in turn, been strongly linked with the trait of conscientiousness.<sup>18</sup> Recent studies show that conscientiousness manifests in certain cell phone behaviors. As Stachl et al. (2020) found:

“[H]igher scores [for conscientiousness] were in general predicted for higher mean usage numbers of weather apps, longer usage durations of a local public transportation app [...], longer and less variant usage times of the camera, and less variation in the usage of apps from the checkout and monitoring category. (Stachl et al. 2020, p. 11.)

Furthermore, it turns out that earlier-than-average first and last time usage of cell phones during the day, and increased battery charge time per day are both predictive of conscientiousness. As banks increasingly rely on algorithms that factor in such personality traits for credit predictions, parameters encoding information about the use of weather and transportation apps, about camera use, and charging duration thus suddenly become important for evaluating an individual’s financial perspective.

But again, it is easy to appreciate why we may struggle to relate these kinds of alternative data parameters with ordinary representations of us as financial individuals. If an individual is denied a loan based on their infrequent use of weather and transportation apps, or their infrequent and

<sup>16</sup> See, for instance, Hiller and Jones (2022).

<sup>17</sup> See, for instance, Peltonen et al. (2020) and Stachl et al. (2020).

<sup>18</sup> For the link between delayed gratification and increased credit score, see, for instance, Meier and Sprenger (2010). For the link between delayed gratification and conscientiousness, see Furnham and Cheng (2019).

unsteady charging of their phones, it is clear that they can feel misrepresented or even misunderstood as a financial being by the algorithm. After all, the path from metrics like phone charge durations to creditworthiness is indirect and complex at best, and it is not something that we would expect individuals to consider when they inquire about a loan and aim to demonstrate sensible financial behavior. As we also saw above, the algorithmic representations of us as certain types of statistical individuals are too disconnected from how we ordinarily understand and represent ourselves as individuals.

While there are many other examples to give of algorithmic representations that appeal to such alternative or non-standard parameters, we trust our main point is clear: when machine learning algorithms make predictions about individuals, they often rely on properties that need not directly relate to nor directly reflect the individuals who are affected by these algorithmic predictions. Let us call such properties *non-self-relatable properties*. What counts as a non-self-relatable property depends on the individual in question and the context of the algorithmic prediction. For a typical individual, for instance, the property of leaving one's cell phone in the charger overnight will count as a non-self-relatable property in the context of the individual's finances and creditworthiness. Likewise, for a typical individual, the property of having a certain number of social workers involved in one's job placement case will count as a non-self-relatable property in the context of assessing the individual's job readiness. As a very rough criterion for whether a property P should count as non-self-relatable to an individual with respect to a particular context C, we can imagine that the individual asks herself whether P should be among the properties mentioned in the parts of her completed autobiography or life story that unfolds in C. If the answer is "no", P counts as non-self-relatable to the individual with respect to C. If the answer is "yes", P counts as self-relatable to the individual with respect to C.<sup>19</sup> For example, when reviewing the comprehensive story of his life, Peter might question himself whether the property of leaving one's cell phone in the charger overnight should be mentioned in the sections dedicated to his financial journey. Presumably, Peter would think 'not', in which case details about phone charge durations would be absent from—and neither clearly implied by—the financial parts of his life story. In this sense, properties reflecting facts about phone

charge duration will count as non-self-relatable for Peter in the context of his financial life.<sup>20</sup>

Of course, we have not attempted to give a precise definition of what constitutes a non-self-relatable property, but in light of the examples above, we trust that the characterization is informative enough to work with. As motivated above, complex algorithmic predictions will likely appeal to many properties that will count as non-self-relatable to the individuals that are affected by these predictions. Accordingly, we can say that when machine learning algorithms represent individuals, they represent individuals as statistical individuals: as statically defined bundles of self-relatable and non-self-relatable properties. In contrast to biological, psychological, and narrative properties, as we have seen, non-self-relatable properties are very different from the kinds of properties that we ordinarily take as being relevant for settling questions about what makes us the specific individuals that we are.

## 5 Concluding remarks

Algorithmic predictions hold promise of assisting—if not improving—human decision making across a range of crucial societal sectors such as banking, administration, and healthcare. To fully capitalize on the power of these predictions, however, we must be ready to be treated as statistical individuals. That is, we must be ready to accept that algorithms can impact us as actual people based on their representations of us as statistically determined combinations of self-relatable and non-self-relatable properties. In part, as we have seen, this means accepting that there need be no biological, psychological or narrative continuity between the individual affected by an algorithm prediction and the statistical individual represented by the algorithm for purposes of said prediction. In part, this also means recognizing that attributes such as phone charging habits and weather app usage—seemingly unrelated to our financial

<sup>19</sup> If there are cases where it is unobvious whether a property is relatable or non-self-relatable to an individual, we may also talk about properties that are neither relatable nor non-self-relatable. Since what matters for our purposes are mainly the non-self-relatable properties, we shall not dwell further on these issues here.

<sup>20</sup> Admittedly, as pointed out by a referee, the criterion for self-relatability above is demanding, if taken too literally. For it may be inscrutable whether a specific property P belongs to one's completed autobiography or not. Yet, we need not interpret the criterion too literally. While we may not know all the details of our life stories, certain properties intuitively seem extremely unlikely to warrant a space in an autobiography. For example, one may reasonably exclude cell phone charger behavior, although this exclusion could be reconsidered if it is discovered to be crucial for an individual's financial identity. This demonstrates that what counts as a self-relatable property can change over time, just as what is included in one's autobiography can change over time. In general, though, the criterion for self-relatability only serves as a rough heuristics, and there are many complicated issues left to consider about how to reconcile aspects of narrative identity with algorithmic representations.

predicament—can affect facets of our lives like which financial prospects and opportunities we have.

Of course, as we saw in the discussion of the narrative view, we are already accustomed to having many different factors influence our representations of individuality. While we may all agree that families and societies play a constitutive role in *unifying* the temporally extended stories of our lives, the high-dimensional vector representations that machine learning algorithms utilize do not have this effect. As we have seen, since many of the properties that define a statistical individual are non-self-relatable, there is no guarantee that they will play a unifying role in an individual's life story. If anything, the fact that machine learning algorithms utilize non-self-relatable properties for making predictions about us can feel fractionating and alienating to our self-conceptions of who we are as individuals. This is of course not to say that we cannot in principle expand our understanding and representation of human identity to include such non-self-relatable properties. But even then, the algorithmic representation of us as statistical individuals will yield a revisionary conception of individuality that it will take us time and cognitive effort to feel—if possible at all—ownership and responsibility for.

Our main aim in this paper was to unpack how machine learning algorithms in a data-rich reality represent human identity. We have argued that the statisticalized representation of individuals differs significantly from our ordinary conception of human identity. Algorithmic representations give no special attention to biological, psychological, and narrative properties, and, as such, they fail to capture central aspects of our ordinary representation of human identity. Indeed, we have argued, algorithms make predictions about us by relying on properties that do not directly relate to nor reflect the individuals we understand and represent ourselves as being. However, we have only touched the tip of the iceberg, and more work is needed to discern how these algorithmic representations affect our ordinary self-conception and sense of responsibility, our social and ethical practices, and our existential outlook.<sup>21</sup>

To give a more concrete example: our conclusions seem straightforwardly relevant for the ever-growing literature on algorithmic fairness and justice. The literature on algorithmic fairness attempts to find purely statistical criteria for what makes an algorithm fair towards different groups of people.<sup>22</sup> For instance, a predictive algorithm may count as unfair if it yields an unequal amount of false positives and

false negatives for people because of their race, gender, or sexual orientation. The literature on algorithmic justice aims to lay down different criteria for what makes an algorithm just. An algorithm, for instance, may count as unjust if it discriminates against certain population groups based on their race or gender, or if it ignores certain structural features of social inequality between different groups.<sup>23</sup> These ethical issues are well documented, and they have a common source in the specific groups that individuals belong to. That is, whether an algorithm counts as unfair or unjust depends, in some sense, on its treatment of individuals who relevantly differ solely based on their membership in groups defined by properties such as race, gender, or sexual orientation. But if we are right, algorithm representations of human identity may constitute another potential ground of algorithmic unfairness and injustice. For if algorithmic representations can conflict with our own representations of who we are, they can also seemingly conflict with our sense of responsibility. Suppose, for instance, that an algorithm predicts that an individual X is not eligible for a loan because some future, purely statisticalized version of X defaults on the loan—we may assume that the statisticalized version of X stands in no biological, psychological, or narrative continuity with the actual individual X. In such a case, X may well feel unfairly and unjustly treated by the algorithm because X does not feel any ownership of and responsibility for the statisticalized representation of him. Accordingly, the discussions in this paper can contribute a new ethical perspective to the ongoing debates on algorithmic fairness and justice.<sup>24</sup>

**Author contributions** All the authors contributed to idea development, analysis, and writing.

**Funding** Open access funding provided by Aarhus Universitet. This work was generously supported by a *Carlsberg Foundation Young Researcher Fellowship (Grant Number CF20-0257)* to Jens Christian Bjerring.

**Data availability** Not applicable.

**Materials and/or code availability** Not applicable.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Ethics approval** Not applicable.

<sup>21</sup> For some recent discussions that it may be both epistemically and morally problematic to use purely statistical evidence as basis for making (algorithmic) decisions about individuals, see Bolinger (2021), Holm (2023a), and Holm (2023b).

<sup>22</sup> While there are many entry points to the discussions of algorithmic fairness, Hedden (2021) is a good place to start.

<sup>23</sup> See Himmelreich (2023) and Marjanovic et al. (2022) for some recent, fairly general discussions of algorithmic justice.

<sup>24</sup> We wish to thank two anonymous reviewers for this journal for their constructive and helpful comments, which have significantly improved the clarity of the paper.

**Consent to participate** Not applicable.

**Consent to publish** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ammitzbøll Flügge A, Holten Møller N, Hildebrandt T, Palmer Olsen H (2022) Er du grøn—algoritmer til beslutningsstøtte i det offentlige. En kvalitativ undersøgelse af sagsbehandleres praksis og brug af ASTA til profilering af nyledige dagpengemodtagere. (Are you green—algorithms for decision support in the public sector. A qualitative study of case handlers' practice and use of ASTA for profiling newly unemployed unemployment benefit recipients). Department of Computer Science, University of Copenhagen. <https://static1.squarespace.com/static/5e3ad7fa73600c394b539f6b/t/628e3f91ba738054d6b7d90c/1653489554932/Er+du+gr%C3%B8n+Algoritmer+til+beslutningsst%C3%B8tte+%28F1%C3%BCgge+et+al.%2C+2022%29.pdf>. Accessed 7 Dec 2023
- Ayodele TO (2010) Types of machine learning algorithms. *New Adv Mach Learn* 3:19–48
- Babushkina D, Votsis A (2022) Disruption, technology and the question of (artificial) identity. *AI Ethics* 2(4):611–622
- Bjerring JC, Busch J (2021) Artificial intelligence and patient-centered decision-making. *Philos Technol* 34(2):349–371
- Bolinger R (2021) Explaining the justificatory asymmetry between statistical and individualized evidence. In: Robson J, Hoskins Z (eds) *The social epistemology of legal trials*. Routledge, pp 60–76
- Clark A, Chalmers D (1998) The extended mind. *Analysis* 58(1):7–19
- de Vries K (2010) Identity, profiling algorithms and a world of ambient intelligence. *Ethics Inf Technol* 12:71–85
- DeGrazia D (2005) *Human identity and bioethics*. Cambridge University Press, Cambridge
- del Valle JI, Lara F (2023) AI-powered recommender systems and the preservation of personal autonomy. *AI Soc* 1–13
- Furnham A, Cheng H (2019) The big-five personality factors, mental health, and social-demographic indicators as independent predictors of gratification delay. *Personal Individ Differ* 150:109533
- Garnelo M, Shanahan M (2019) Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Curr Opin Behav Sci* 29:17–23
- Hedden B (2021) On statistical criteria of algorithmic fairness. *Philos Public Aff* 49(2):209–231
- Hiller JS, Jones LS (2022) Who's keeping score?: oversight of changing consumer credit infrastructure. *Am Bus Law J* 59(1):61–121
- Himmelreich J (2023) Against “democratizing AI.” *AI Soc* 38(4):1333–1346
- Holm S (2023a) Data-driven decisions about individual patients: the case of medical AI. *J Eval Clin Pract*
- Holm S (2023b) Statistical evidence and algorithmic decision-making. *Synthese* 202(1):28
- Humphreys MS, Yoshihisa K (2002) Connectionism and self: distributed representational systems and their implications for self and identity. In: *Self and identity*. Psychology Press, pp 27–54
- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2(4)
- Kharif O (2016) No credit history? no problem. lenders are looking at your phone data. [Bloomberg.com](https://www.bloomberg.com)
- Kripke S (1972) *Naming and necessity*. Harvard University Press, Cambridge
- Lagerkvist A, Tudor M, Smolicki J, Ess CM, Eriksson Lundström J, Rogg M (2022) Body stakes: an existential ethics of care in living with biometrics and AI. *AI Soc* 1–13
- LenndoEFL (n.d.) <https://lendoefl.com>. Accessed 25 Oct 2022
- Leo M, Sharma S, Maddulety K (2019) Machine learning in banking risk management: a literature review. *Risks* 7(1):29
- Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Denniston AK (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health* 1(6):e271–e297
- Marjanovic O, Cecez-Kecmanovic D, Vidgen R (2022) Theorising algorithmic justice. *Eur J Inf Syst* 31(3):269–287
- Meier S, Sprenger C (2010) Present-biased preferences and credit card borrowing. *Am Econ J Appl Econ* 2(1):193–210
- Milano S, Taddeo M, Floridi L (2020) Recommender systems and their ethical challenges. *AI Soc* 35:957–967
- Olson ET (1999) *The human animal: personal identity without psychology*. Oxford University Press, Oxford
- Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B (2019) The value of big data for credit scoring: enhancing financial inclusion using mobile phone data and social network analytics. *Appl Soft Comput* 74:26–39
- Parfit D (1986) *Comments*. *Ethics* 96:832–872
- Pedersen NJLL, Bjerring JC (2022) Extended knowledge overextended? In: *Knowers and knowledge in East-West philosophy: epistemology extended*, pp 191–233
- Peltonen E, Sharmila P, Asare KO, Visuri A, Lagerspetz E, Ferreira D (2020) When phones get personal: predicting big five personality traits from application usage. *Pervasive Mob Comput* 69:101269
- Petersen AC, Cohn ML, Hildebrandt T, Møller NH (2021) Thinking problematically’ as a resource for AI design in politicised contexts. In: *CHIItaly 2021: 14th biannual conference of the Italian SIGCHI chapter*, pp 1–8
- Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2(3):160
- Schechtman M (1996) *The constitution of selves*. Cornell University Press, Ithaca
- Schechtman M (2011) The narrative self. In: Gallagher S (ed) *The oxford handbook of the self*. Oxford University Press, Oxford, pp 394–416
- Schechtman M (2014) *Staying alive: personal identity, practical concerns, and the unity of a life*. OUP, Oxford
- Seidelin C, Moreau T, Shklovski I, Holten Møller N (2022) Auditing risk prediction of long-term unemployment. In: *Proceedings of the ACM on human-computer interaction*, 6(GROUP), pp 1–12
- Shoemaker S (1963) *Self-knowledge and self-identity*. Cornell University Press, Ithaca

- Shoemaker S (2008) Personal identity and ethics: a brief introduction. Broadview Press, Ontario
- Shoemaker D (2021) Personal identity and ethics. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy (Fall 2021 Edition). <https://plato.edu/archives/fall2021/entries/identity-ethics/>
- Søe SO (2021) Nonnatural personal information. Accounting for misleading and non-misleading personal information. *Philos Technol* 34(4):1243–1262
- Søe SO, Mai J (2022) Data identity: privacy and the construction of self. *Synthese* 200(6):492
- Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, Bühner M (2020) Predicting personality from patterns of behavior collected with smartphones. *Proc Natl Acad Sci* 117(30):17680–17687
- Sutton J (1998) *Philosophy and memory traces: descartes to connectionism*. Cambridge University Press, Cambridge

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.